

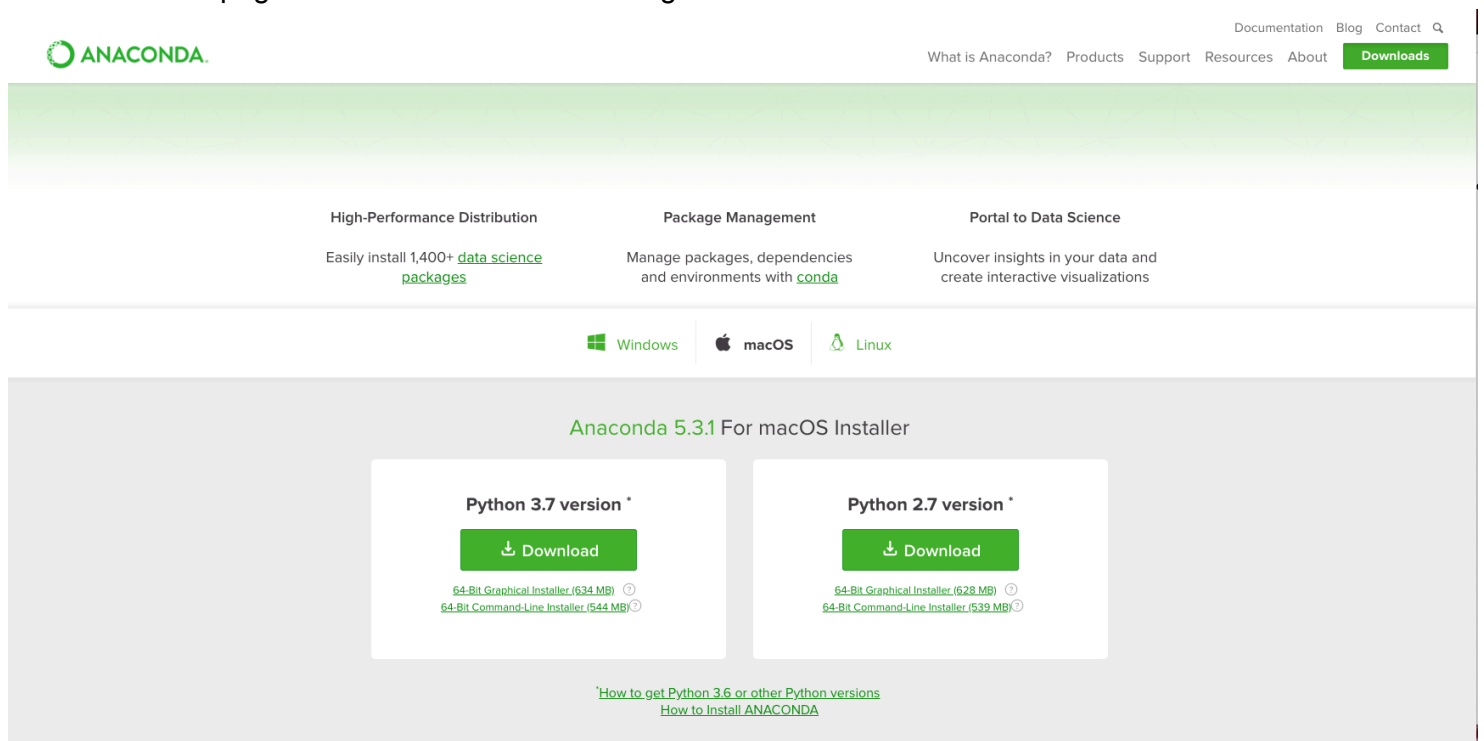
Instruction Manual

Step 1. Install Anaconda

Anaconda is a software that allows you to run the Jupyter Notebook on your computer. Please refer to this link for download:

<https://www.anaconda.com/download/>

Scroll down the page a little bit to find the following section:



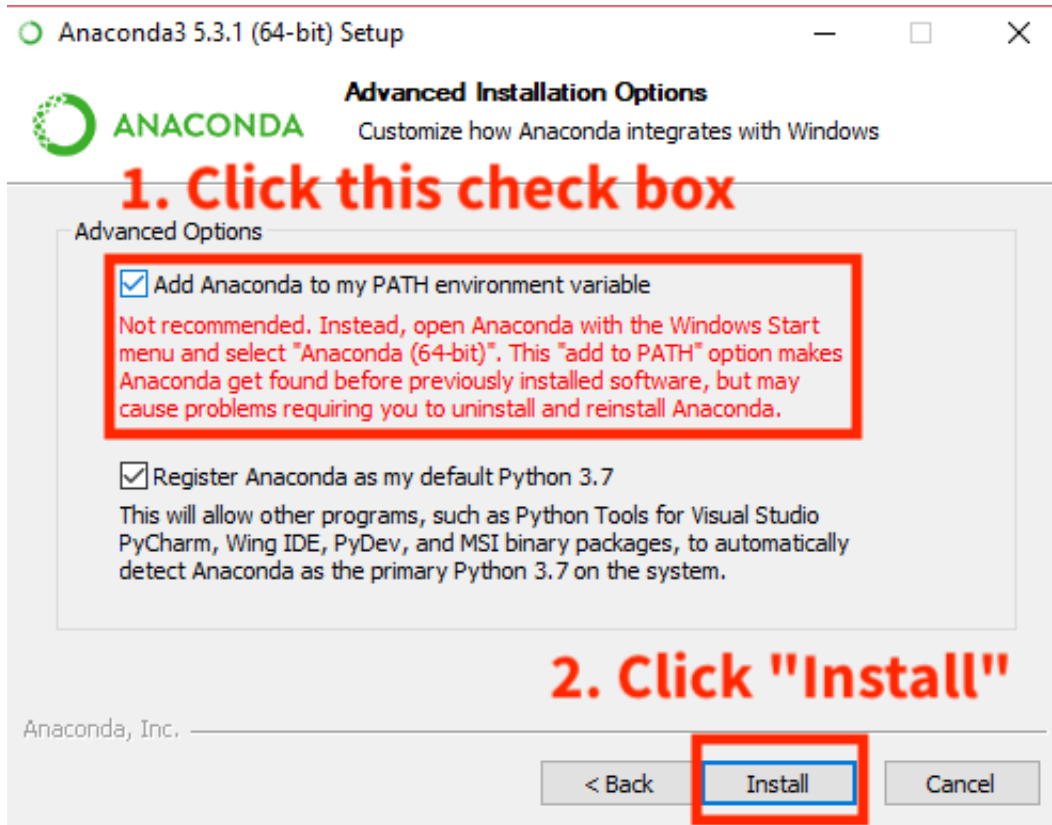
The screenshot shows the Anaconda website's download section. At the top, there is a navigation bar with links for Documentation, Blog, Contact, and a search icon. Below this, there are three main sections: High-Performance Distribution, Package Management, and Portal to Data Science. The Package Management section is highlighted, showing that it manages packages, dependencies, and environments with conda. Below these sections, there are icons for Windows, macOS, and Linux. The macOS section is selected, showing the Anaconda 5.3.1 For macOS Installer. There are two download options: Python 3.7 version and Python 2.7 version. Each option has a green 'Download' button and links to 64-bit Graphical and Command-Line installers. At the bottom, there are links for 'How to get Python 3.6 or other Python versions' and 'How to Install ANACONDA'.

Please download the **Python 3.7 version** of Anaconda

Then please install Anaconda using the downloaded executable.

Windows User

Please execute the installer and keep click on '**Next**' without changing any settings until you see this page:



Make sure you check the option **"Add Anaconda to my PATH environment variable"** before you click **"Install"**.

If the installer is asking you to install "MS Visual Studio Code", please ignore it.

Step 2. Install Required Python Libraries

We are now installing the required Python libraries to allow the program to execute correctly.

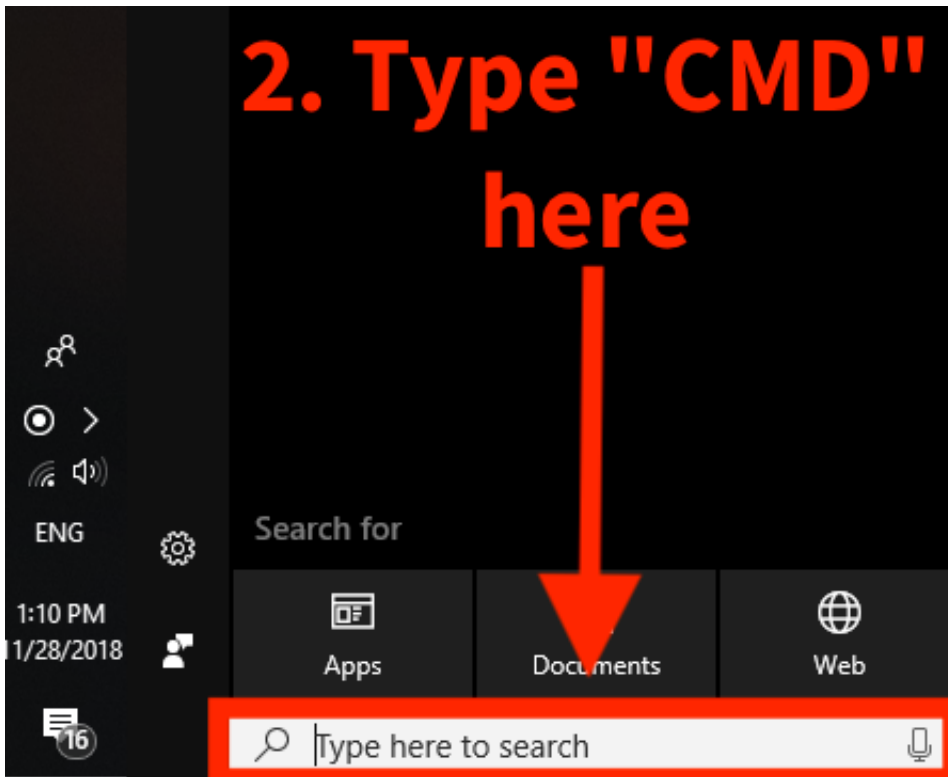
Windows User

1. Please open **CMD** program with "Administrator Privilege"

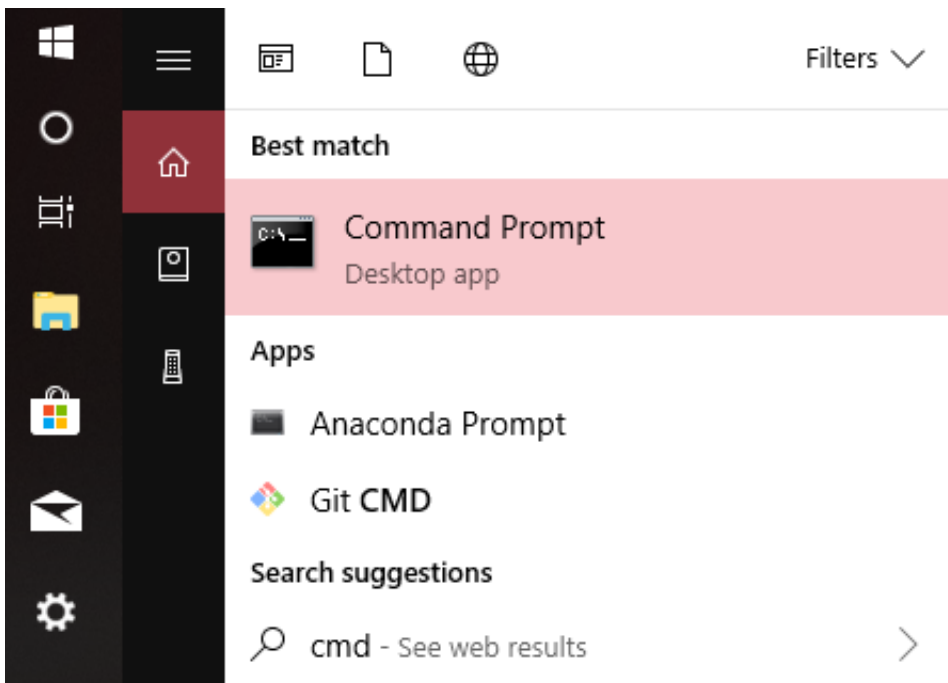
On the Windows taskbar, find a circle icon and click it



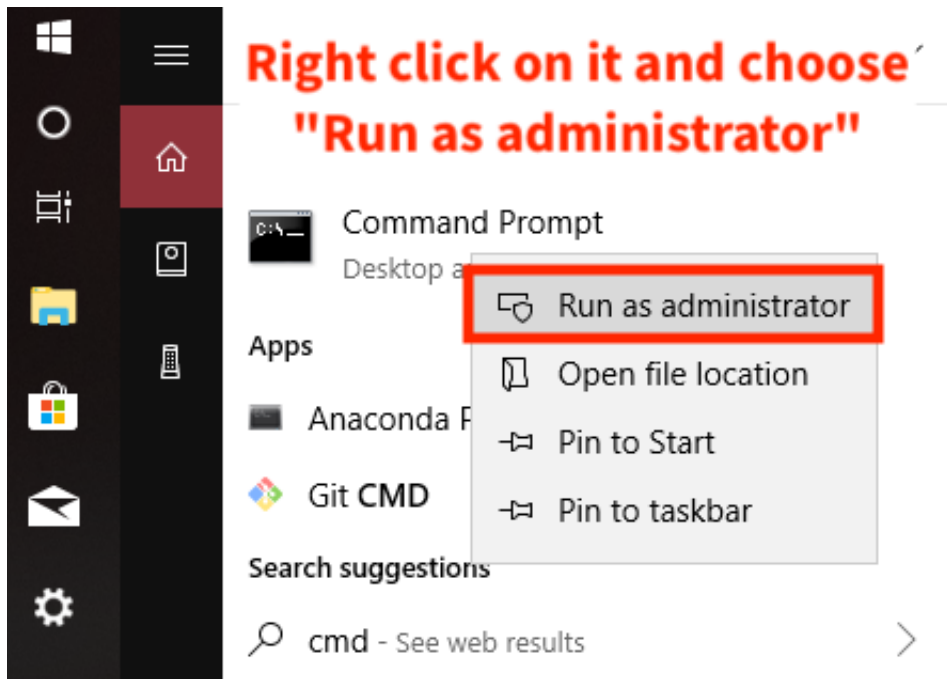
You will see a input field at the button, type "cmd"



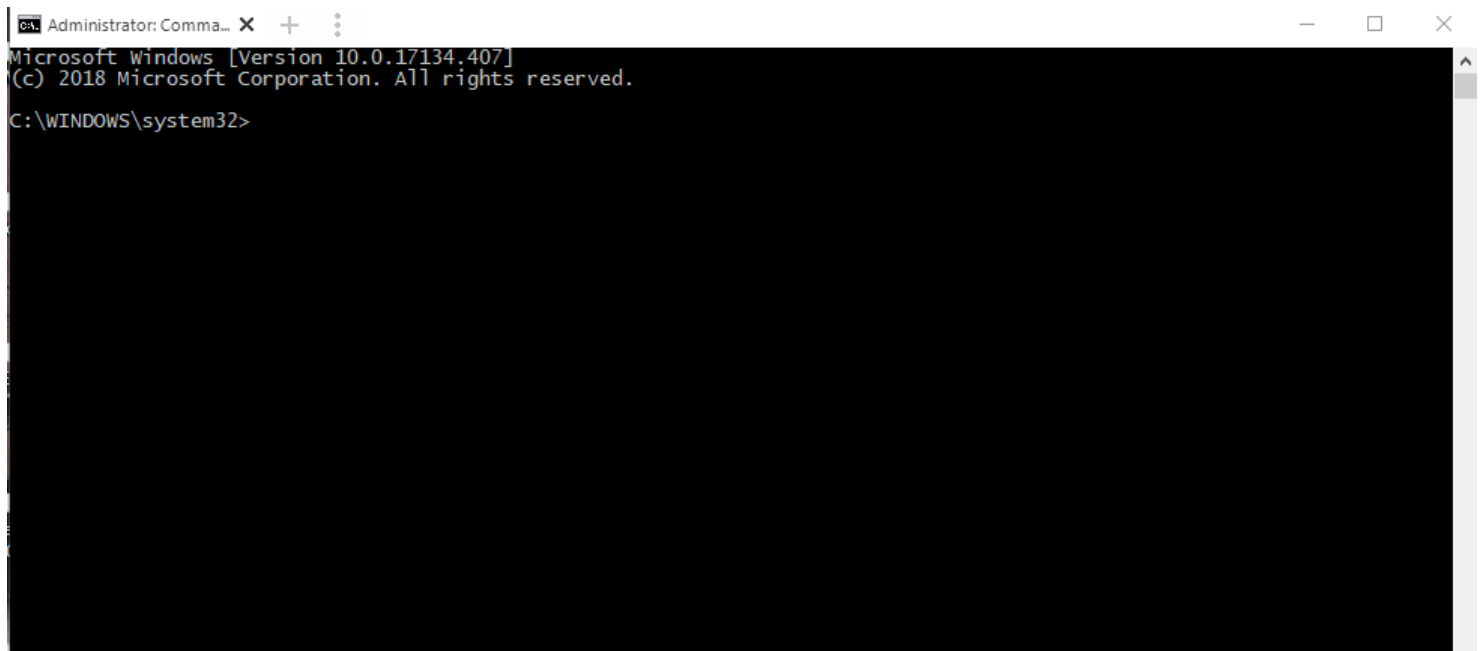
The cmd program will show up in the window.



Please right click on it and choose "Run as administrator"



Here is the CMD window:



2. Insert the command to install the libraries:

This is the command for installation:

```
conda install -c conda-forge selenium pandas numpy
```

Copy & paste this line to the CMD window and press "Enter" key.

```
Administrator: Comma... X + ...
Microsoft Windows [Version 10.0.17134.407]
(c) 2018 Microsoft Corporation. All rights reserved.
C:\WINDOWS\system32>conda install -c conda-forge selenium pandas numpy

Copy & Paste the command here and press "Enter" key
```

Then, you will see a message like the following, type 'y' and press "Enter" key.

```
The following NEW packages will be INSTALLED:
  libflang: 5.0.0-vc14_20180208 conda-forge [vc14]
  llvm-meta: 5.0.0-0 conda-forge
  openblas: 0.3.3-h535eed3_1001 conda-forge
  openmp: 5.0.0-vc14_1 conda-forge [vc14]

The following packages will be UPDATED:
  blas: 1.0-mkl --> 1.1-openblas conda-forge
  numpy: 1.15.1-py37ha559c80_0 --> 1.15.4-py37_bblas_openblash442142e_1000 conda-forge [blas_openblas]
  pandas: 0.23.4-py37h830ac7b_0 --> 0.23.4-py37h830ac7b_1000 conda-forge

Proceed ([y]/n)? y Type 'y' and press "Enter" key
```

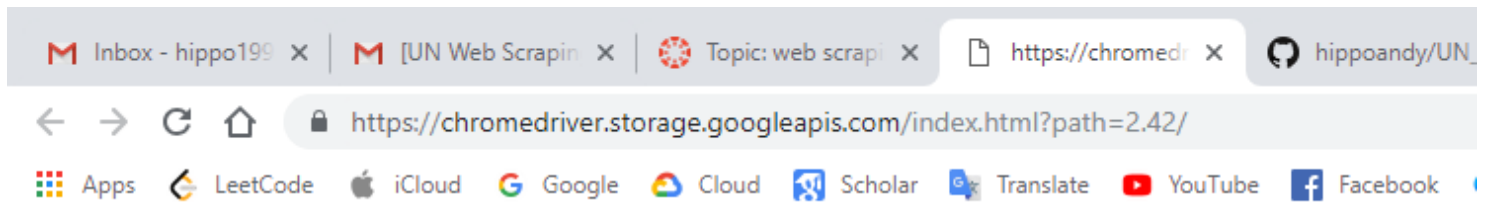
Wait until the progress complete.

Step 3. Download additional Lib. file and the Program

1. Download Chrome Driver

Please refer to this link: <https://chromedriver.storage.googleapis.com/index.html?path=2.42/>

Download Windows version Chrome Driver as:

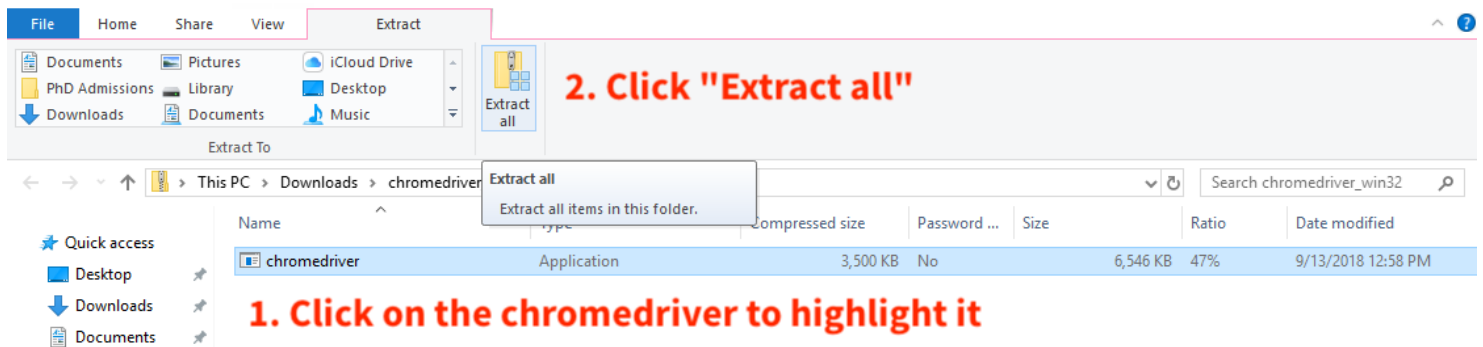


Index of /2.42/

Name	Last modified	Size	ETag
Parent Directory			
chromedriver_linux64.zip	2018-09-13 19:30:37	3.85MB	acfcc29fb03df9e913ef4c360a121ad1
chromedriver_mac64.zip	2018-09-13 18:14:11	5.75MB	3fc0e4a97cbf2c8c2a9b824d95e25351
chromedriver_win32.zip	2018-09-13 21:11:33	3.42MB	28d91b31311146250e7ef1afbcd6d026
notes.txt	2018-09-13 21:23:09	0.02MB	18bdf6fc9f9d8dd668fa444b77d06bdd

Click this one

Open the downloaded .zip file and extract the "chromedriver.exe" to your **Downloads** folder



You will see a window shows up, change the path to your **Download** folder and click "**Extract**"



← Extract Compressed (Zipped) Folders

Select a Destination and Extract Files

Files will be extracted to this folder:

C:\Users\Andy\Downloads\

Browse...

Show extracted files when complete

1. change the path to the "Downloads" folder

2. Click "Extract"

Extract

Cancel

2. Download the program files:

Please refer to this link: <https://github.com/hippoandy/UNWeb scrapingWORKANA/>

Download the program .zip file as:

hippoandy / UN_Web scraping_WORKANA

40 commits 1 branch

Clone or download

Download ZIP

1. Click "Clone or download"

2. Click "Download ZIP"

Open the downloaded .zip file and extract the content:

UN_Web scraping_WORKANA-mas... File folder

Extract all

Click "Extract all"

Make sure to change to destination path to your **Download** folder:



← Extract Compressed (Zipped) Folders

Select a Destination and Extract Files

Files will be extracted to this folder:

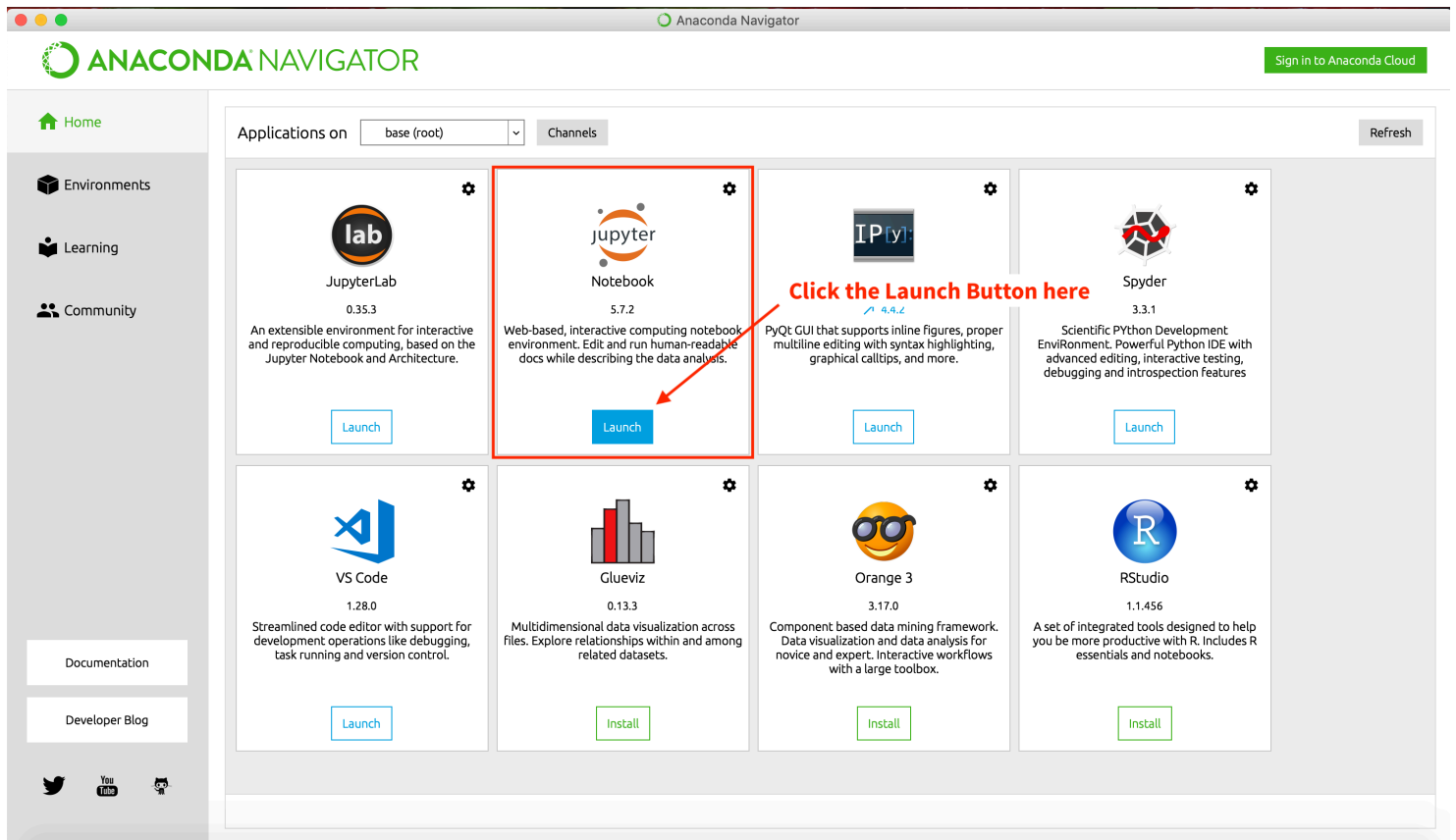
Show extracted files when complete

**Make sure to change the path
to your download folder**

Step 4. Open Jupyter Notebook Application

1. Open Anaconda then Jupyter

Open you installed Anaconda Program, you should see a window like this:



To launch the Jupyter Notebook, click the "**Launch**" button of it.

Then, you should see you browser opens up and shows a page like this:

The screenshot shows the JupyterLab interface with the file browser open to the root directory. The 'Downloads' folder is highlighted with a red box, and a red arrow points to it with the text 'Click it to navigate to your Download folder'.

Name	Last Modified	File size	
0	/		
3D Objects	13 days ago		
Anaconda3	an hour ago		
Contacts	13 days ago		
Desktop	4 days ago		
Documents	2 hours ago		
Downloads	seconds ago		
Dropbox	13 days ago		
Favorites	11 days ago		
iCloudDrive	18 hours ago		
Links	13 days ago		
Music	13 days ago		
OneDrive	18 hours ago		
Pictures	12 days ago		
Roaming	14 days ago		
Saved Games	13 days ago		
Searches	13 days ago		
Videos	12 days ago		

Navigate to your Download folder, and you should see this:

The screenshot shows the JupyterLab interface with the file browser open to the 'Downloads' folder. The 'UN_Webscraping_WORKANA-master' folder is highlighted with a red box, and a red arrow points to it with the text 'Continue to navigate to the program folder'.

Name	Last Modified	File size
0	/ Downloads	
..	seconds ago	
UN_Webscraping_WORKANA-master	9 minutes ago	
chromedriver.exe	an hour ago	6.7 MB

Here you will see the downloaded files instructed in the previous section. Continue to navigate to the program folder:

```
UN_Webscraping_WORKANA-master > scrpaing > selenium
```

Then, you should see this page:

Select items to perform actions on them.

Here is the path, make sure it is the same as yours

Upload New ↕

/ Downloads / UN_Web scraping_WORKANA-master / scraping / selenium		Name ↓	Last Modified	File size
..			seconds ago	
chrome-driver			11 minutes ago	
WORKANA_Freelancers_Scraping_using_Selenium.ipynb			11 minutes ago	15.4 kB
result.csv			11 minutes ago	2 kB
scrap.py			11 minutes ago	5.4 kB
tutorial.py			11 minutes ago	2.11 kB

Open this Jupyter Notebook

Click on the Jupyter Notebook to open the program.

Here is the Jupyter Notebook program

Run the notebook through this button

```
In [1]: '''
This program scraps the information from the freelancers section of the WORKANA website.
This program utilizes the Selenium library for web scraping, which enables the program to change conditions or
filters the website provided to show different result

Author: Yu-Chang (Andy) Ho
Date: 2018/11/27
'''

Out[1]: '\nThis program scraps the information from the freelancers section of the WORKANA website.\nThis program utilizes the Selenium
library for web scraping, which enables the program to change conditions or\nfilters the website provided to show different res
ult\n\nAuthor: Yu-Chang (Andy) Ho\nDate: 2018/11/27\n'
```

```
In [2]: # import the required Libraries
from selenium import webdriver
# the library to control program exection at run time
import time
import pandas as pd
import numpy as np

print( 'This section import the required libraries.' )

This section import the required libraries.
```

```
In [3]: # parameter settings -----

# the keyword which to perform a search
# as an example, we are scraping the freelancers who is familiar with SQL Language
KEY_search = 'SQL'

# the target URL
URL_target = 'https://www.workana.com/freelancers'

# time to wait for the page to refresh
# to make sure the changes applied to the website and allow the page to reload
TIME_pending = 2

# path of the output file
PATH_output = './result.csv'

# the search results will be seperated into multiple subpages
# this is a large number to cover all the page number
limit = 3

# ----- parameter settings

print( 'Here are the parameters that able to be modified.' )

Here are the parameters that able to be modified.
```

```
In [4]: # self-defined functions -----

# return the string 'N/A'
def invalid_val(): return 'N/A'

# make sure there is no special char in a value
def clear_str( text ):
    text = str(text).replace( '\n', '' ).replace( '\r', '' ).replace( '\t', '' )
    return text
```

Please use the run button to execute the notebook.

2. Run the Notebook

Before execution, please find the driver section of the code, the section starts with the line:

```
### load the chrome driver executable
```

Replace the tag to your computer user name:

```
In [6]: ### Load the chrome driver executable

# path to the driver executable

### Windows
PATH_chrome_driver = 'c:/Users/<user>/Downloads/chromedriver.exe'

### macOS
#PATH_chrome_driver = './chrome-driver/chromedriver'

# Load the driver
driver = webdriver.Chrome( PATH_chrome_driver )

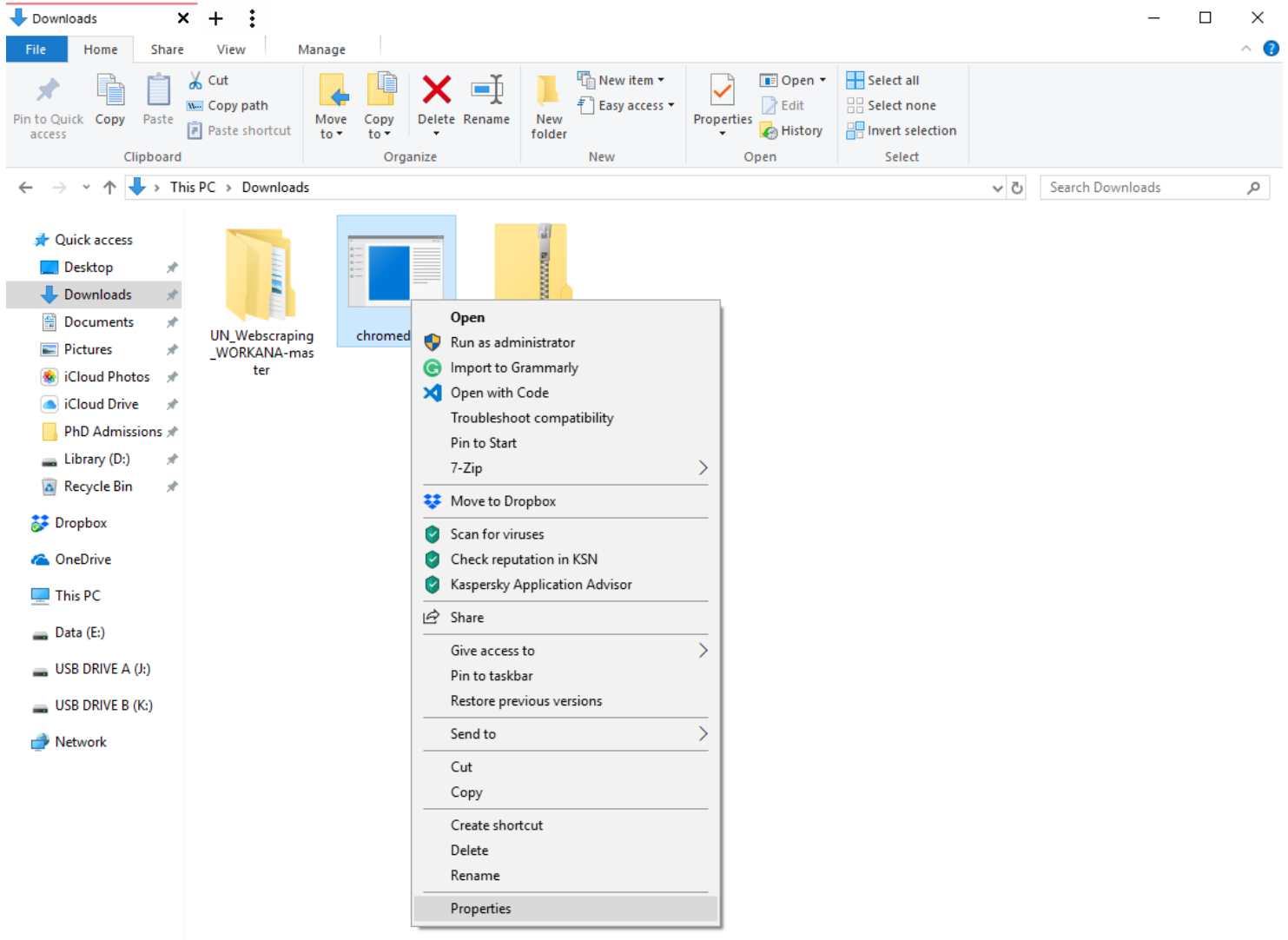
print( 'Loading the chrome driver executable file.' )
print( 'This will open up a blank Chrome browser window.' )
```

Replace '<user>' with your user name

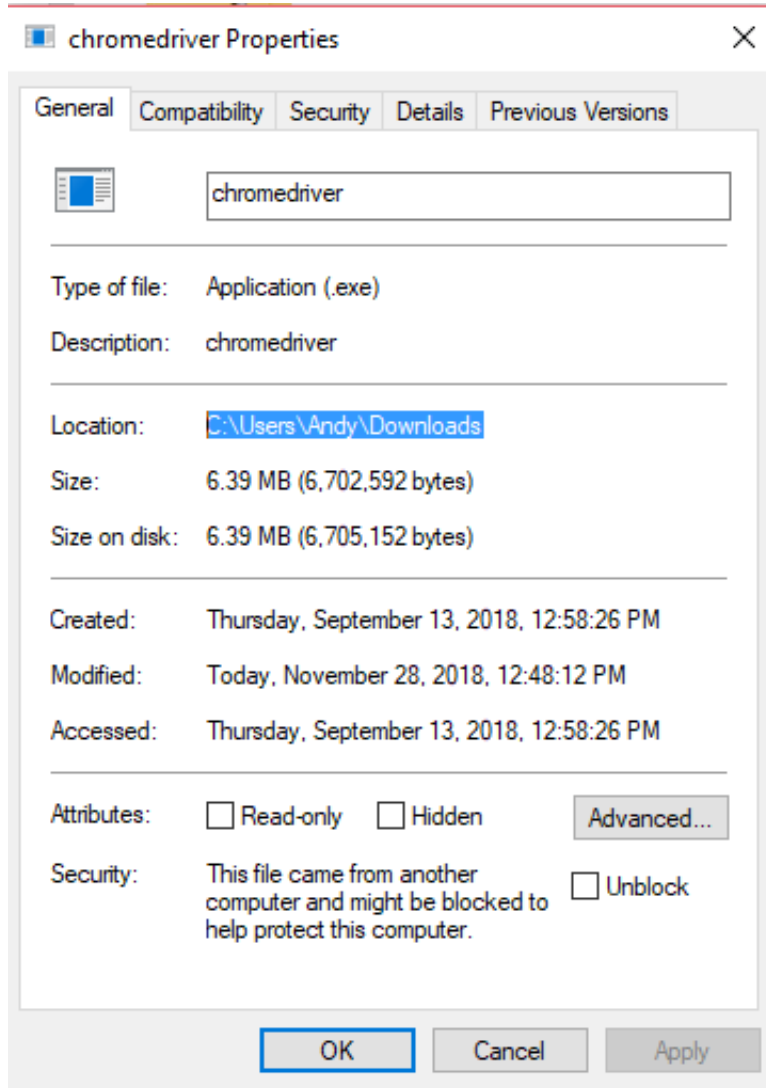
Loading the chrome driver executable file.
This will open up a blank Chrome browser window.

To find your load name, you may check the properties information of the **chromedriver.exe** file.

Headed to your Downloads folder, right click on the **chromedriver.exe** and choose "**Properties**".



You will see a window like this, the location field indicated the user name:



Here, the user name is "**Andy**".

Please go back to the Jupyter notebook and replace the user name with yours.

If you see this window while running the program, please click "**Allow**".



Windows Defender Firewall has blocked some features of this app

Windows Defender Firewall has blocked some features of Python on all public and private networks.



Name: Python
Publisher: Python Software Foundation
Path: C:\users\andy\anaconda3\python.exe

Allow Python to communicate on these networks:

- Private networks, such as my home or work network
- Public networks, such as those in airports and coffee shops (not recommended because these networks often have little or no security)

[What are the risks of allowing an app through a firewall?](#)

Allow access

Cancel